

Estimating Minimum Sum-rate for Cooperative Data Exchange

Ni Ding*, Rodney A. Kennedy* and Parastoo Sadeghi*

*The Research School of Engineering, College of Engineering and Computer Science, the Australian National University (ANU), Canberra, ACT 2601

Email: {ni.ding, rodney.kennedy, parastoo.sadeghi}@anu.edu.au

Abstract—This paper considers how to accurately estimate the minimum sum-rate so as to reduce the complexity of solving cooperative data exchange (CDE) problems. The CDE system contains a number of geographically close clients who send packets to help the others recover an entire packet set. The minimum sum-rate is the minimum value of total number of transmissions that achieves universal recovery (the situation when all the clients recover the whole packet set). Based on a necessary and sufficient condition for a supermodular base polyhedron to be nonempty, we show that the minimum sum-rate for a CDE system can be determined by a maximization over all possible partitions of the client set. Due to the high complexity of solving this maximization problem, we propose a deterministic algorithm to approximate a lower bound on the minimum sum-rate. We show by experiments that this lower bound is much tighter than those lower bounds derived in the existing literature. We also show that the deterministic algorithm prevents from repetitively running the existing algorithms for solving CDE problems so that the overall complexity can be reduced accordingly.

I. INTRODUCTION

Due to the growing amount of data exchange over wireless networks and increasing number of mobile clients, the base-station-to-peer (B2P) links are severely overloaded. It is called the ‘last mile’ bottleneck problem in wireless transmissions. Cooperative peer-to-peer (P2P) communications is proposed for solving this problem. The idea is to allow mobile clients to exchange information with each other through P2P links instead of solely relying on the B2P transmissions. If the clients are geographically close to each other, the P2P transmissions could be more reliable and faster than B2P ones.

Consider the situation when a base station wants to deliver a set of packets to a group of clients. Denote $\mathcal{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_L\}$ the packet set and $\mathcal{K} = \{1, \dots, K\}$ the client set. Due to the fading effects of wireless channels, after several broadcasts via B2P links, there may still exist some clients that do not obtain all the packets. Fig. 1 shows an example when the base station wants to disseminate 6 packets to 3 clients. In this figure, \mathcal{H}_j is the has-set that denotes the packets obtained by client j after B2P transmissions. Since the clients’ knowledge of the packet set may be complementary to each other, instead of relying on retransmissions from the base station, the clients can broadcast combinations of the packets they know via P2P wireless links so as to help the others recover the missing packets. We call Fig. 1 cooperative data exchange (CDE) system. For this kind of systems, there is a so-called CDE problem: how to find an efficient transmission strategy that

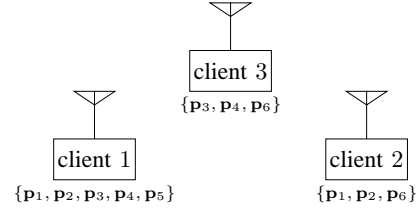


Fig. 1. An example of CDE system: There are three clients that want to obtain six packets. The has-sets are $\mathcal{H}_1 = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_5\}$, $\mathcal{H}_2 = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_6\}$ and $\mathcal{H}_3 = \{\mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_6\}$.

achieves the *universal recovery* (the situation when all clients recover the entire packet set).

Let $\mathbf{r} = (r_1, \dots, r_K)$ be a transmission strategy, where r_j denotes the total number of linear combinations transmitted by client j . The CDE problem is usually expressed in the general form of [1]

$$\begin{aligned} & \min_{\mathbf{r}} f(\mathbf{r}) \\ & \text{s.t. } \sum_{j \in \mathcal{X}} r_j \geq \left| \bigcap_{j \in \mathcal{K} \setminus \mathcal{X}} \mathcal{H}_j^c \right|, \forall \mathcal{X} \subset \mathcal{K}, \\ & \sum_{j \in \mathcal{K}} r_j = \alpha, \end{aligned} \quad (1)$$

where α denotes the transmission budget, the designated total number of transmissions among the clients. In (1), the expression of $f(\mathbf{r})$ is determined based on what kind of strategy is considered to be efficient. For example, if $f(\mathbf{r}) = \mathbf{w}^\top \mathbf{r}$, the most efficient strategy is the one that minimizes the weighted-sum of transmissions [1]; if $f(\mathbf{r}) = \sum_{j \in \mathcal{K}} r_j \log(r_j)$, the most efficient strategy is the one that distribute the transmission rate in the fairest way [2]. For most of the algorithms that solve a particular CDE problem, the value of α is assumed to be known a priori. For example, the algorithms proposed in [1], [2] find minimum weighted sum-rate and fairest strategies, respectively, for a given value of α .

However, problem (1) does not have solutions for all values of α . In fact, there exists a minimum sum-rate α^* so that the constraint set in (1) is nonempty only if $\alpha \geq \alpha^*$. But, the value of α^* is usually unknown in advance. On the other hand, the algorithms proposed in [1], [2] can check the feasibility of a given value of α (i.e., whether there exists a strategy that can achieve universal recovery under budget α). Therefore, for the minimization problem over the minimum sum-rate

strategy set (problem (1) when $\alpha = \alpha^*$), one can start with any arbitrary value of α and adjust it to the feasible values accordingly, e.g., setting α as the lower bound on α^* derived in [3], [4], run the algorithms in [1], [2] and increase the value of α until a feasible solution is found. By doing so, the algorithms in [1], [2] should be called for many times until the final solution is found. But, the complexity of running algorithms proposed in [1], [2] is not low. These algorithms involve submodular function minimizations, the complexity of which is at least $O(K^5\gamma + K^6)$.¹ For this reason, estimating a value of minimum sum-rate that is close to α^* prevents from repetitively running algorithms proposed in [1], [2] and, therefore, helps reduce the overall complexity of solving problem (1).

In this paper, we study how to accurately estimate the minimum sum-rate α^* for both CDE systems that allow packet splitting (PS-CDE) and CDE systems that do not allow packet splitting (NPS-CDE). We start the study with an existing result derived in [1], [2]: the crossing submodularity of $|\bigcap_{j \in \mathcal{K} \setminus \mathcal{X}} \mathcal{H}_j^c|$. By using a necessary and sufficient condition given in [6] for the constraint set in (1) to be nonempty, we show that the exact value of α^* can be determined by a maximization over all possible partitions of the client set \mathcal{K} , which is NP-hard [7]. Instead of calculating the exact value of α^* , we propose a deterministic algorithm to estimate a lower bound on α^* . We show by experiments that the lower bound found by the deterministic algorithm is much tighter than the ones derived in [3], [4]. In fact, the experiment results show that the deterministic algorithm returns the exact value of minimum sum-rate in most cases. In this paper, we also discuss how our results are related to those in [1], [3], [4], [8] and prove that any PS-CDE problem can be converted to an NPS-CDE problem.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Let $\mathcal{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_L\}$ be the packet set containing L linearly independent packets. Each packet \mathbf{p}_i belongs to a field \mathbb{F}_q . The system contains K geographically close clients. Define the client set as $\mathcal{K} = \{1, \dots, K\}$. Each client $j \in \mathcal{K}$ initially obtains $\mathcal{H}_j \subset \mathcal{P}$. Here, \mathcal{H}_j is called the *has-set* of client j . We also denote $\mathcal{H}_j^c = \mathcal{P} \setminus \mathcal{H}_j$ as the packet set that is missing at client j . The clients are assumed to collectively know the the packet set, i.e., $\bigcup_{j \in \mathcal{K}} \mathcal{H}_j = \mathcal{P}$. The P2P wireless links between clients are error-free, i.e., any information broadcast by client j can be heard losslessly by client j' for all $j' \in \mathcal{K} \setminus \{j\}$. The clients broadcast linear combinations of the packets in their has-sets in order to help each other recover the entire packet set \mathcal{P} . For example, in the CDE system in Fig. 1, client 1 broadcasting $\mathbf{p}_1 + \mathbf{p}_3$ helps client 2 recover \mathbf{p}_3 and client 3 recover \mathbf{p}_1 , and client 2 broadcasting $\mathbf{p}_1 + \mathbf{p}_6$ helps client 1 recover \mathbf{p}_6 and client 3 recover \mathbf{p}_1 .

For a transmission strategy $\mathbf{r} = (r_1, \dots, r_K)$, where r_j denotes the total number of linear combinations transmitted by

client j , we call $\sum_{j \in \mathcal{K}} r_j$ the sum-rate of strategy \mathbf{r} . Denote \mathcal{R}_α the set that contains all transmission strategies that achieve universal recovery and have sum-rate equal to constant α . \mathcal{R}_α is nonempty only if α is greater or equal to the minimum sum-rate α^* . For example, consider the CDE system in Fig. 1 and assume it is an NPS-CDE system. Since $\mathbf{r} \in \mathbb{N}_0^K$, $\alpha \in \mathbb{N}_0$, i.e., we only consider nonnegative integer values of α . In this system, $\alpha \leq 3$, $\mathcal{R}_\alpha = \emptyset$, i.e., there is no strategy with a sum-rate less than or equal to 3 that can achieve universal recovery. But, if $\alpha = 4$, there exist three strategies that achieve universal recovery: $\mathcal{R}_4 = \{(3, 0, 1), (3, 1, 0), (2, 1, 1)\}$. For example, in strategy $(3, 0, 1)$, client 1 broadcasting $\mathbf{p}_1 + \mathbf{p}_3$, $\mathbf{p}_2 + \mathbf{p}_4$ and \mathbf{p}_5 and client 3 broadcasting \mathbf{p}_6 can achieve the universal recovery. Therefore, the minimum sum-rate is $\alpha^* = 4$. Assume Fig. 1 is a PS-CDE system. Since $\mathbf{r} \in \mathbb{R}_+^K$, $\alpha \in \mathbb{R}_+$, i.e., we consider the nonnegative real values of α . It can be shown that $\alpha^* = 3.5$ with the corresponding strategy set being $\mathcal{R}_{3.5} = \{(2.5, 0.5, 0.5)\}$ and $\mathcal{R}_\alpha = \emptyset$ if $\alpha < 3.5$. Note, the way to implement strategy $(2.5, 0.5, 0.5)$ is to break each packet into two chunks, e.g., the has-set of client 3 would be $\{\mathbf{p}_3^{(1)}, \mathbf{p}_3^{(2)}, \mathbf{p}_4^{(1)}, \mathbf{p}_4^{(2)}, \mathbf{p}_6^{(1)}, \mathbf{p}_6^{(2)}\}$. Letting client 1 transmit 5 chunks and both client 2 and client 3 transmit 1 chunk is sufficient to achieve the universal recovery. In this case, the packet normalized strategy is $(2.5, 0.5, 0.5)$.

It is shown in [1] that the CDE problems can be formulated by (1) in general. Note, (1) is an integer programming problem for NPS-CDE systems. It can be seen from the above examples that for any CDE system there exist a minimum sum-rate α^* such that \mathcal{R}_α is nonempty only if $\alpha \geq \alpha^*$, i.e., problem (1) does not have a solution when $\alpha < \alpha^*$ since the constraint set is empty. Therefore, estimating the value of the minimum sum-rate α^* helps determine whether an α is feasible or not, and, as discussed before, reduce the overall complexity of solving problem (1).

III. MINIMUM SUM-RATE

As discussed in Section II, estimating the minimum sum-rate α^* is related to the nonemptiness of \mathcal{R}_α . In this section, we show a method to determine α^* by studying the condition for the nonemptiness of \mathcal{R}_α . We use a necessary and sufficient condition given in [6] for \mathcal{R}_α to be nonempty to show how to determine α^* .

A. Constant Sum-rate Set

Let $\mathbf{r}(\mathcal{X}) = \sum_{j \in \mathcal{X}} r_j$ and define

$$g(\mathcal{X}) = \left| \bigcap_{j \in \mathcal{K} \setminus \mathcal{X}} \mathcal{H}_j^c \right|. \quad (2)$$

It is proved in [1] that a transmission strategy \mathbf{r} can achieve universal recovery if

$$\mathbf{r}(\mathcal{X}) \geq g(\mathcal{X}) \quad (3)$$

for all \mathcal{X} such that $\mathcal{X} \subset \mathcal{K}$. The interpretation of (3) is: the information sent from \mathcal{X} should be complement to that missing in $\mathcal{K} \setminus \mathcal{X}$, i.e., the total number of packets transmitted by the clients in any set \mathcal{X} should be greater or equal to the number of packets that are commonly missing at the remaining clients.

¹There are many algorithms proposed for submodular function minimization problem. To our knowledge, the algorithm proposed in [5] has the lowest complexity $O(K^5\gamma + K^6)$, where γ is the complexity of evaluating a submodular function.

We can describe the constraint set in problem (1) as the constant sum-rate set

$$\mathcal{R}_\alpha = \left\{ \mathbf{r} \in \mathbb{R}_+^K : \mathbf{r}(\mathcal{X}) \geq g(\mathcal{X}), \forall \mathcal{X} \subset \mathcal{K}, \mathbf{r}(\mathcal{K}) = \alpha \right\} \quad (4)$$

for a PS-CDE system. For an NPS-CDE system, $\mathbf{r} \in \mathbb{R}_+^K$ should be replaced with $\mathbf{r} \in \mathbb{N}_0^K$. Therefore, determining minimum sum-rate is equivalent to finding the smallest value of α such that \mathcal{R}_α is nonempty.

B. Determining Minimum Sum-rate

It is proved in Lemma 1 in [2] and Lemma 6 in [1] that g is a crossing supermodular function.² Let $\mathcal{I} \subseteq \{1, 2, \dots\}$ with $2 \leq |\mathcal{I}| \leq K$ and denote $\{\mathcal{X}_i\}_{i \in \mathcal{I}}$ a partition of \mathcal{K} .³ We can use the following proposition to determine whether \mathcal{R}_α is empty or not.

Proposition 3.1 (nonemptiness of \mathcal{R}_α): \mathcal{R}_α is nonempty if and only if

$$\alpha \geq \sum_{i \in \mathcal{I}} g(\mathcal{X}_i), \quad (5)$$

$$\alpha \leq \sum_{i \in \mathcal{I}} \left(\alpha - g(\mathcal{K} \setminus \mathcal{X}_i) \right) \quad (6)$$

for all partitions $\{\mathcal{X}_i\}_{i \in \mathcal{I}}$.

Proof: This is a direct result of Theorem 2.6 in [6], which derives the necessary and sufficient condition for a crossing supermodular base polyhedron to be nonempty. The crossing supermodular base polyhedron in our case is \mathcal{R}_α . ■

Based on Proposition 3.1, we can determine the minimum sum-rate α^* by the following theorem.

Theorem 3.2: The minimum sum-rate is

$$\alpha^* = \max \left\{ \sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1} : \text{all partitions } \{\mathcal{X}_i\}_{i \in \mathcal{I}} \right\}. \quad (7)$$

for PS-CDE systems and

$$\alpha^* = \max \left\{ \left\lceil \sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1} \right\rceil : \text{all partitions } \{\mathcal{X}_i\}_{i \in \mathcal{I}} \right\} \quad (8)$$

for NPS-CDE systems, where $\lceil y \rceil$ is the ceiling function which returns the minimum integer value that is no less than y .

Proof: Eq (6) in Proposition 3.1 can be rewritten as

$$\alpha \geq \sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1}. \quad (9)$$

But, since $g(\mathcal{X} \cup \mathcal{Y}) \geq g(\mathcal{X}) + g(\mathcal{Y})$ for all $\emptyset \neq \mathcal{X}, \mathcal{Y} \subset \mathcal{K}$, it can be shown that

$$\sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1} \geq \sum_{i \in \mathcal{I}} g(\mathcal{X}_i).^4 \quad (10)$$

So, \mathcal{R}_α is nonempty if and only if $\alpha \geq \sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1}$ for all partitions. Therefore, for a PS-CDE system, the minimum

²A set function $f: 2^\mathcal{K} \mapsto \mathbb{R}_+$ is supermodular if for all $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{K}$ if $f(\mathcal{X}) + f(\mathcal{Y}) \leq f(\mathcal{X} \cap \mathcal{Y}) + f(\mathcal{X} \cup \mathcal{Y})$ for all $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{K}$ such that $\mathcal{X} \cap \mathcal{Y} \neq \emptyset$, $\mathcal{X} - \mathcal{Y} \neq \emptyset$, $\mathcal{Y} - \mathcal{X} \neq \emptyset$ and $\mathcal{X} \cup \mathcal{Y} \neq \mathcal{K}$ [6].

³A partition $\{\mathcal{X}_i\}_{i \in \mathcal{I}}$ of \mathcal{K} satisfies $\mathcal{X}_i \neq \emptyset$, $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ and $\cup_{i \in \mathcal{I}} \mathcal{X}_i = \mathcal{K}$ for all $i, j \in \mathcal{I}$.

⁴Please see the proof in Appendix A.

sum-rate is determined by (7). For an NPS-CDE system, α^* must be the smallest integer that is greater or equal to $\sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1}$ for all partitions, which can be expressed by (8). ■

Consider what (5) and (9) mean in the CDE system. Assume that the clients can form groups, or coalitions, under the condition that any client can only appear in at most one coalition. Then, any form of coalition can be represented by $\{\mathcal{X}_i\}_{i \in \mathcal{I}}$, a partition of \mathcal{K} . Based on $\{\mathcal{X}_i\}_{i \in \mathcal{I}}$, consider determining the value of α such that \mathcal{R}_α is nonempty by using (3). On one hand, $\mathbf{r}(\mathcal{X}_i) \geq g(\mathcal{X}_i)$ for all $i \in \mathcal{I}$. Recall that $\alpha = \mathbf{r}(\mathcal{K}) = \sum_{i \in \mathcal{I}} \mathbf{r}(\mathcal{X}_i)$. We have the condition $\alpha \geq \sum_{i \in \mathcal{I}} g(\mathcal{X}_i)$, which is exactly (5). On the other hand, $\mathbf{r}(\mathcal{K} \setminus \mathcal{X}_i) \geq g(\mathcal{K} \setminus \mathcal{X}_i)$ for all $i \in \mathcal{I}$. We have $\sum_{i \in \mathcal{I}} \mathbf{r}(\mathcal{K} \setminus \mathcal{X}_i) \geq \sum_{i \in \mathcal{I}} g(\mathcal{K} \setminus \mathcal{X}_i)$. It is equivalent to $(|\mathcal{I}| - 1)\alpha \geq \sum_{i \in \mathcal{I}} g(\mathcal{K} \setminus \mathcal{X}_i)$, which is exactly (9). Since (3) should be satisfied for all subsets that are not equal to \mathcal{K} , (5) and (9) should be satisfied for all partitions. But, according to (10), (9) is more strict than (5). So, satisfying (9) for all partitions is sufficient to determine α^* as stated in Theorem 3.2.

Example 3.3: Consider the CDE system in Fig. 1. We have function g as

$$\begin{aligned} g(\emptyset) &= 0, g(\{1\}) = 1, g(\{2\}) = 0, g(\{3\}) = 0, \\ g(\{1, 2\}) &= 3, g(\{1, 3\}) = 3, g(\{2, 3\}) = 1. \end{aligned}$$

By applying Theorem 3.2, we have $\alpha^* = 3.5$ for PS-CDE system and $\alpha^* = 4$ for NPS-CDE system. The corresponding strategy sets are $\mathcal{R}_{3.5} = \{(2.5, 0.5, 0.5)\}$ and $\mathcal{R}_4 = \{(2, 1, 1), (3, 0, 1), (3, 1, 0)\}$. It can be shown that \mathcal{R}_α is empty for all $\alpha < 3.5$.

C. Relationship with Existing Works

In [8], it was shown that the minimum sum-rate is determined by

$$\alpha^* = L - \min \left\{ \frac{\sum_{i \in \mathcal{I}} |\bigcup_{j \in \mathcal{X}_i} \mathcal{H}_j| - L}{|\mathcal{I}| - 1} : \text{all partitions } \{\mathcal{X}_i\}_{i \in \mathcal{I}} \right\}.^5 \quad (11)$$

One can show that it is exactly (7). Alternatively speaking, the authors in [8] derive the the minimum sum-rate for PS-CDE systems.

For PS-CDE systems, Theorem 3 in [1] states that dividing packet into $K - 1$ chunks is sufficient to achieve the normalized minimum sum-rate with high probability. But, Theorem 3.2 in this paper establishes that dividing each packet into $K - 1$ chunks is sufficient to achieve the normalized minimum sum-rate for sure. In addition, define

$$\hat{g} = (K - 1)g(\mathcal{X}). \quad (12)$$

We can also conclude from Theorem 3.2 that there exists a minimum integer valued α such that

$$\hat{\mathcal{R}}_\alpha = \left\{ \mathbf{r} \in \mathbb{N}_0^K : \mathbf{r}(\mathcal{X}) \geq \hat{g}(\mathcal{X}), \forall \mathcal{X} \subset \mathcal{K}, \mathbf{r}(\mathcal{K}) = \alpha \right\} \quad (13)$$

⁵It is found in [9] that the solutions of CDE and secrecy generation problems can be found by solving the same linear programming (LP) problem. Eq (11) was originally proposed in [10] to solve a secrecy generation problem. It is also a method to determine the minimum sum-rate of CDE systems [8].

Algorithm 1: Deterministic Algorithm

Initiate the sum-rate lower bound: $\beta = \lceil \sum_{j \in \mathcal{K}} \frac{g(\mathcal{K} \setminus \{j\})}{K-1} \rceil$;
for $k = 1$ **to** K **do**
 $\mathcal{W}_1 = \{k\}$;
 for $m = 2$ **to** $K - 1$ **do**
 $u_m^* = \arg \max \{g(\mathcal{K} \setminus (\mathcal{W}_{m-1} \cup \{u\})) - g(\mathcal{K} \setminus \{u\}) : u \in \mathcal{K} \setminus \mathcal{W}_{m-1}\}$;
 $\mathcal{W}_m = \mathcal{W}_{m-1} \cup \{u_m^*\}$;
 $\beta = \max\{\beta, \lceil \frac{g(\mathcal{K} \setminus \mathcal{W}_m) + \sum_{j \in \mathcal{K} \setminus \mathcal{W}_m} g(\mathcal{K} \setminus \{j\})}{|\mathcal{K} \setminus \mathcal{W}_m|} \rceil\}$;
 endfor
endfor
Output β ;

is nonempty, and that value is the non-normalized minimum sum-rate for the PS-CDE system. Therefore, any PS-CDE problem can be converted to NPS-CDE one (See the example below). For this reason, in the rest of this paper, we consider NPS-CDE problems only.

Example 3.4: Consider the CDE system in Fig. 1 and assume it is a PS-CDE system. Let each packet be divided into $K - 1 = 2$ chunks. We replace g with \hat{g} as

$$\begin{aligned} \hat{g}(\emptyset) &= 0, \hat{g}(\{1\}) = 2, \hat{g}(\{2\}) = 0, \hat{g}(\{3\}) = 0, \\ \hat{g}(\{1, 2\}) &= 6, \hat{g}(\{1, 3\}) = 6, \hat{g}(\{2, 3\}) = 2. \end{aligned}$$

By using (8) in Theorem 3.2, we get $\alpha^* = 7$. Therefore, the normalized minimum sum-rate is $7/2 = 3.5$, which is in consistence with the result in Example 3.3.

IV. TIGHT LOWER BOUND ON MINIMUM SUM-RATE

Problem (8) is equivalent to a minimum k -partition problem. The NP-hardness of this problem is proved in [7], i.e., to directly determine the minimum sum-rate by Theorem 3.2 is intractable. Therefore, we propose a deterministic algorithm as shown in Algorithm 1 to approximate α^* . In this algorithm, we initiate β , the estimation of the minimum sum-rate, as the value of $\lceil \sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}|-1} \rceil$ in the case of K -partition. We then update β by considering the value of $\lceil \sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}|-1} \rceil$ under a set of m -partitions where $2 \leq m \leq K - 1$. From the following theorem, we show that the output β of Algorithm 1 is a lower bound on α^* .

Theorem 4.1: For any CDE system, the output β of Algorithm 1 satisfies $\beta \leq \alpha^*$.

Proof: We remark that \mathcal{W}_m in Algorithm 1 for all $2 \leq m \leq K - 1$ is generated based on Queyranne's algorithm proposed in [11]. Here, \mathcal{W}_m and all $j \in \mathcal{K} \setminus \mathcal{W}_m$ form a $|\mathcal{K} \setminus \mathcal{W}_m| + 1$ -partition that contains $|\mathcal{K} \setminus \mathcal{W}_m|$ singletons. It is proved in [11] that

$$g(\mathcal{K} \setminus \mathcal{W}_m) + g(\mathcal{K} \setminus \{j\}) \geq g(\mathcal{K} \setminus (\mathcal{W}_m \setminus \mathcal{X})) + g(\mathcal{K} \setminus (\mathcal{X} \cup \{j\})),$$

for all $j \in \mathcal{K} \setminus \mathcal{W}_m$, $\mathcal{X} \subseteq \mathcal{W}_{m-1}$ and $2 \leq m \leq K - 1$. Therefore,

$$\begin{aligned} &g(\mathcal{K} \setminus \mathcal{W}_m) + \sum_{j \in \mathcal{K} \setminus \mathcal{W}_m} g(\mathcal{K} \setminus \{j\}) \\ &\geq g(\mathcal{K} \setminus (\mathcal{W}_m \setminus \mathcal{X})) + g(\mathcal{K} \setminus (\{j'\} \cup \mathcal{X})) \\ &\quad + \sum_{j \in \mathcal{K} \setminus (\mathcal{W}_m \cup \{j'\})} g(\mathcal{K} \setminus \{j\}), \end{aligned}$$

i.e., $\lceil \frac{g(\mathcal{K} \setminus \mathcal{W}_m) + \sum_{j \in \mathcal{K} \setminus \mathcal{W}_m} g(\mathcal{K} \setminus \{j\})}{|\mathcal{K} \setminus \mathcal{W}_m|} \rceil$ is the maximum value of $\lceil \sum_{i \in \mathcal{X}_i} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}|-1} \rceil$ over a subset of $|\mathcal{K} \setminus \mathcal{W}_m| + 1$ -partitions for all m . Therefore, based on Theorem 3.2, the output β of Algorithm 1 is a lower bound on minimum sum-rate α^* . ■

A. Relationship with Existing Works

In Lemma 2 in [3], it is shown that the minimum sum-rate is lower bounded as

$$\alpha^* \geq \max_{j \in \mathcal{K}} |\mathcal{H}_j^c| = \max_{j \in \mathcal{K}} g(\mathcal{K} \setminus \{j\}). \quad (14)$$

In Lemma 10 in [4], it is shown that the minimum sum-rate is lower bounded as

$$\alpha^* \geq \left\lceil \frac{\sum_{j \in \mathcal{K}} |\mathcal{H}_j^c|}{K-1} \right\rceil = \left\lceil \frac{\sum_{j \in \mathcal{K}} g(\mathcal{K} \setminus \{j\})}{K-1} \right\rceil. \quad (15)$$

But, it can be seen that the output β of Algorithm 1 satisfies

$$\beta \geq \max \left\{ \max_{j \in \mathcal{K}} \{g(\mathcal{K} \setminus \{j\}) + g(\{j\})\}, \left\lceil \frac{\sum_{j \in \mathcal{K}} g(\mathcal{K} \setminus \{j\})}{K-1} \right\rceil \right\}.$$

Since $\max_{j \in \mathcal{K}} \{g(\mathcal{K} \setminus \{j\}) + g(\{j\})\} \geq \max_{j \in \mathcal{K}} g(\mathcal{K} \setminus \{j\})$, β is tighter than the lower bounds given in [3], [4]. To see how tight β is, we run the following experiments.

Example 4.2: We vary the number of clients K from 3 to 15 and the number of packets L from 6 to 30. For each combination of K and L , we repeat the procedure below for 1000 times.

- randomly generate the has-sets \mathcal{H}_j for all $j \in \mathcal{K}$ subject to the condition $\cup_{j \in \mathcal{K}} \mathcal{H}_j = \mathcal{P}$;
- get the minimum sum-rate α^* by using the randomized algorithm proposed in [4]; calculate the lower bounds given in [3] and [4]; obtain the lower bound β by running Algorithm 1.

We take the error as the absolute value of the difference between the lower bound and the minimum sum-rate α^* and the average error as the mean error over 1000 repetitions. We show the average error of lower bounds in [3], [4] and Algorithm 1 in Figs. 2, 3 and 4, respectively. It can be seen that the lower bound found by Algorithm 1 is much tighter than the ones in [3], [4] and, in most cases, Algorithm 1 finds the exact value of α^* .

B. Complexity

It should be noted that running Algorithm 1 alone is not able to completely solve problem (1). To find a solution of problem (1) and also check if the universal recovery can be achieved with the sum-rate $\alpha = \beta$, one needs to run the algorithms proposed in [1], [2]. But, running Algorithm 1

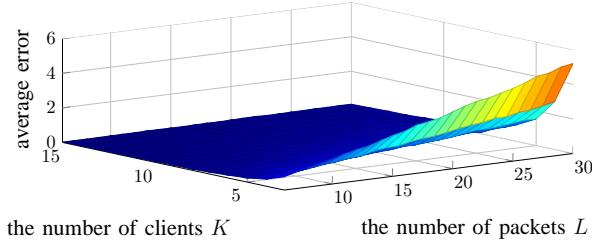


Fig. 2. The average error of lower bound in [3] when K varies from 3 to 15 and L varies from 6 to 30.

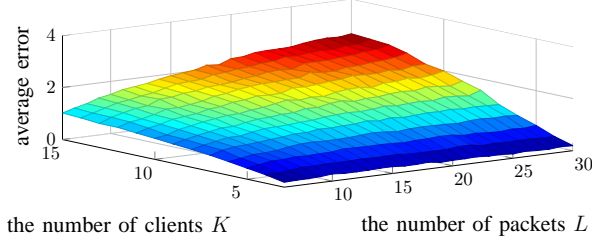


Fig. 3. The average error of lower bound in [4] when K varies from 3 to 15 and L varies from 6 to 30.

before those algorithms in [1], [2] helps reduce the overall complexity.

As aforementioned, the complexity of algorithms in [1], [2] largely depends on the complexity of solving an submodular function minimization problem. For example, the complexity of the algorithm in [1] is $O(K \cdot SFM(K))$, where $SFM(K)$ is the complexity of running a submodular function minimization algorithm. Although it is proved that submodular function minimization problem can be solved in polynomial time, the lowest complexity of $SFM(K)$ as proposed in [5] is still $O(K^5\gamma + K^6)$, where γ is the complexity of evaluating function g . On the other hand, the algorithms proposed in [1], [2] also require a certain value of α as an input. For example, the authors in [1] suggested use an initial value of α equal to the lower bound in [3], repeat running an algorithm with complexity $O(K \cdot SFM(K))$ and increase α until a feasible solution is found. Based on Fig. 2, it means that this algorithm should be called for approximately five times before it finds a feasible solution when $K = 3$ and $L = 30$. On the contrary, the complexity of Algorithm 1 is just $O(K^3\gamma)$.⁶ Based on Fig. 4 if we choose the initial value of α to be β , the number of repetitions of running the algorithm in [1], and hence the overall complexity, will be largely reduced. In fact, in most cases, the algorithm in [1] will be called just once (since $\beta = \alpha^*$ in most cases). In Fig. 5, we show the maximum error of β collected in Example 4.2. It can be seen that the maximum error of β is one, which means that if the $\beta \neq \alpha^*$, the number of repetitions of running the algorithm in [1] will not exceed two.

⁶Algorithm 1 contains $K(K-2)$ iterations, and each iteration involves a maximization over client subset $\mathcal{K} \setminus \mathcal{W}_{m-1}$ by evaluating function g . Therefore, the complexity is $O(K^3\gamma)$.

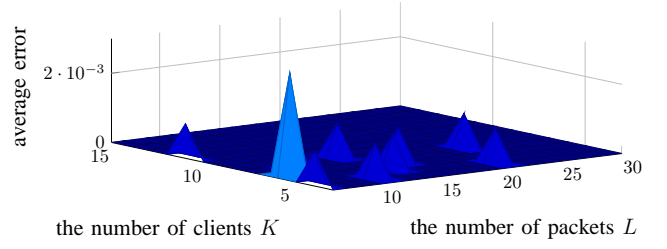


Fig. 4. The average error of lower bound β found by Algorithm 1 when K varies from 3 to 15 and L varies from 6 to 30.

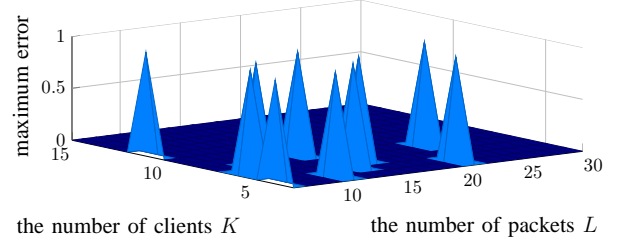


Fig. 5. The maximum error of lower bound β found by Algorithm 1 when K varies from 3 to 15 and L varies from 6 to 30.

V. CONCLUSION

In this paper, we showed that the minimum sum-rate α^* in a CDE system could be determined by a maximization over all possible partitions of the client set. Instead of solving the maximization problem directly, we proposed a deterministic algorithm to estimate the lower bound on α^* and showed by experiment that this lower bound was much tighter than those derived in the existing literature. The experiment results also showed that the output of Algorithm 1 was the exact value of α^* in most cases and the maximum difference between it and α^* is one, which could avoid repetitively running the existing algorithms for solving CDE problems so that the overall complexity could be reduced.

VI. ACKNOWLEDGEMENT

This work is supported under Australian Research Council Discovery Projects funding scheme (project no. DP120100160).

REFERENCES

- [1] T. Courtade and R. Wesel, "Coded cooperative data exchange in multi-hop networks," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1136–1158, Feb. 2014.
- [2] N. Milosavljevic, S. Pawar, M. Gastpar, and K. Ramchandran, "Efficient algorithms for the data exchange problem under fairness constraints," in *Proc. Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, 2012, pp. 502–508.
- [3] S. El Rouayheb, A. Sprintson, and P. Sadeghi, "On coding for cooperative data exchange," in *Proc. IEEE Inform. Theory Workshop*, Cairo, 2010, pp. 1–5.
- [4] A. Sprintson, P. Sadeghi, G. Booker, and S. El Rouayheb, "A randomized algorithm and performance bounds for coded cooperative data exchange," in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, TX, 2010, pp. 1888–1892.
- [5] M. Goemans and V. Ramakrishnan, "Minimizing submodular functions over families of sets," *Combinatorica*, vol. 15, no. 4, pp. 499–513, 1995.
- [6] S. Fujishige, *Submodular functions and optimization*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 2005.

- [7] O. Goldschmidt and D. S. Hochbaum, “A polynomial algorithm for the k-cut problem for fixed k,” *Math. of Oper. Res.*, vol. 19, no. 1, pp. 24–37, 1994.
- [8] N. Milosavljevic, S. Pawar, S. El Rouayheb, M. Gastpar, and K. Ramchandran, “Deterministic algorithm for the cooperative data exchange problem,” in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, 2011, pp. 410–414.
- [9] T. Courtade and R. Wesel, “Weighted universal recovery, practical secrecy, and an efficient algorithm for solving both,” in *Proc. Annu. Allerton Conf. Commun., Control, and Comput.*, Monticello, IL, 2011, pp. 1349–1357.
- [10] C. Chan, “Generating secret in a network,” Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, 2010.
- [11] M. Queyranne, “Minimizing symmetric submodular functions,” *Math. Programming*, vol. 82, no. 1-2, pp. 3–12, Jun. 1998.

Appendices

APPENDIX A

Let $\emptyset \neq \mathcal{X}, \mathcal{Y} \subset \mathcal{K}$ such that $\mathcal{X} \cap \mathcal{Y} = \emptyset$. We have

$$\begin{aligned}
 & g(\mathcal{X} \cup \mathcal{Y}) - g(\mathcal{X}) - g(\mathcal{Y}) \\
 &= \left| \bigcap_{j \in \mathcal{K} \setminus (\mathcal{X} \cup \mathcal{Y})} \mathcal{H}_j^c \right| - \left| \bigcap_{j \in \mathcal{K} \setminus \mathcal{X}} \mathcal{H}_j^c \right| - \left| \bigcap_{j \in \mathcal{K} \setminus \mathcal{Y}} \mathcal{H}_j^c \right| \\
 &= \left| \bigcap_{j \in \mathcal{K} \setminus (\mathcal{X} \cup \mathcal{Y})} \mathcal{H}_j^c \right| - \left| \bigcap_{j \in \mathcal{K} \setminus \mathcal{X}} \mathcal{H}_j^c \cup \bigcap_{j \in \mathcal{K} \setminus \mathcal{Y}} \mathcal{H}_j^c \right| \\
 &= \left| \bigcap_{j \in \mathcal{K} \setminus (\mathcal{X} \cup \mathcal{Y})} \mathcal{H}_j^c \right| - \left| \bigcap_{j \in \mathcal{K} \setminus (\mathcal{X} \cup \mathcal{Y})} \mathcal{H}_j^c \cap \left(\bigcap_{j \in \mathcal{X}} \mathcal{H}_j^c \cup \bigcap_{j \in \mathcal{Y}} \mathcal{H}_j^c \right) \right| \\
 &\geq 0,
 \end{aligned}$$

i.e., $g(\mathcal{X} \cup \mathcal{Y}) \geq g(\mathcal{X}) + g(\mathcal{Y})$.⁷ Then, for any partition $\{\mathcal{X}_i\}_{i \in \mathcal{I}}$ of \mathcal{K} ,

$$g(\mathcal{K} \setminus \mathcal{X}_i) \geq \sum_{i' \in \mathcal{I}, i' \neq i} g(\mathcal{X}_{i'}). \quad (16)$$

So, $\sum_{i \in \mathcal{I}} g(\mathcal{K} \setminus \mathcal{X}_i) \geq (|\mathcal{I}| - 1) \sum_{i \in \mathcal{I}} g(\mathcal{X}_i)$, i.e.,

$$\sum_{i \in \mathcal{I}} \frac{g(\mathcal{K} \setminus \mathcal{X}_i)}{|\mathcal{I}| - 1} \geq \sum_{i \in \mathcal{I}} g(\mathcal{X}_i). \quad (17)$$

⁷Since $g(\emptyset) = 0$, g can be considered supermodular in all \mathcal{X} such that $\emptyset \neq \mathcal{X} \subset \mathcal{K}$.